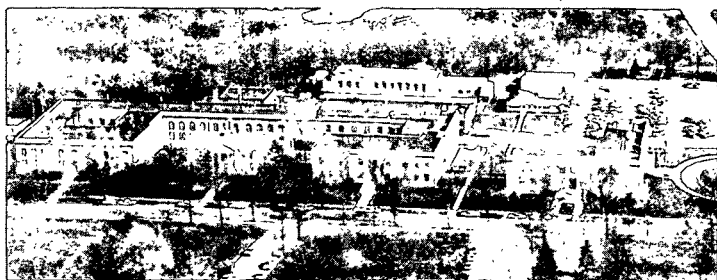


FEB 28 1975  
IPC Tech Pap 2



**THE INSTITUTE OF PAPER CHEMISTRY, APPLETON, WISCONSIN**

**IPC TECHNICAL PAPER SERIES**

**NUMBER 2**

**AUTOMATED POSTING OF BROAD TERMS TO THE KEYWORD SUPPLEMENT  
OF THE ABSTRACT BULLETIN OF THE INSTITUTE  
OF PAPER CHEMISTRY (ABIPC)**

**JOHN O. CHURCH**

**JANUARY, 1975**

AUTOMATED POSTING OF BROAD TERMS TO THE KEYWORD SUPPLEMENT  
OF THE ABSTRACT BULLETIN OF THE INSTITUTE  
OF PAPER CHEMISTRY (ABIPC)

JOHN O. CHURCH

ABSTRACT

The procedure for indexing the Keyword Supplement of the Abstract Bulletin has been expanded. An automatic system for assigning the broad terms for those manually assigned is performed. The additional index terms enable one to take a more general approach to locating abstracts. A comparative test was conducted for twelve standard profiles, and the results are given.

## INTRODUCTION

Beginning with Volume 42 of the Abstract Bulletin of The Institute of Paper Chemistry, indexing of abstracts has been performed using the Thesaurus of Pulp and Paper Terms as the authority. The thesaurus provides a hierarchy for all terms to be used in indexing; narrower, broader, related, used for, or use index terms if any are listed for each permissible term.

## MANUAL INDEXING

Indexers at the Institute manually assign index terms to abstracts to the extent that the essential content of the abstract is indicated. These "keywords" are keypunched and verified, then are processed through the computer system providing the inverted file and direct file of the Keyword Supplement of the Abstract Bulletin in machine-readable and hard copy versions.

## SEMIAUTOMATIC INDEXING

With Volume 44, Number 1 (July, 1974) a new dimension was added to the procedure. After the manual indexing process is completed, the broad terms for keywords assigned are automatically added before producing the magnetic tapes and printed versions. Now someone searching for information, and using the Keyword Supplement, about computer applications only has to look up "COMPUTERS" which is automatically posted for articles which had "DIGITAL COMPUTERS" or "ANALOG COMPUTERS" manually assigned. We say that the Keyword Supplement is upwards general or downwards specific.

## COMPUTERIZED INFORMATION RETRIEVAL

Computer systems have been developed which will search the magnetic tape version of the Keyword Supplement and report back abstracts which had specified combinations of keywords assigned to them. Unless the request is very general or very specific, two groups of index terms are usually organized and abstracts which have a word from each of the groups are retrieved. For instance, if "computer applications in information processing" was the topic of interest, the two groups of words might be

DIGITAL COMPUTERS  
COMPUTER PROGRAMS  
DATA PROCESSING  
AUTOMATION

INFORMATION RETRIEVAL  
INFORMATION SYSTEMS  
DATA RETRIEVAL

All abstracts assigned a word from column one and a word from column two would be retrieved. References which are not pertinent to the topic are sometimes retrieved along with the useful abstracts. Sometimes these "false drops" are more numerous than the "hits." The Institute offers monthly information on twenty-seven standard subject profiles. Twelve of these profiles have been tested to determine the effect of having the broad terms posted. Table I shows the results of the two searches.

TABLE I  
SEARCH OF VOLUME 44, NUMBER 7

| <u>Broad Terms Not Posted</u> |           |             |           | <u>Broad Terms Posted</u> |             |           |            |                   |
|-------------------------------|-----------|-------------|-----------|---------------------------|-------------|-----------|------------|-------------------|
| Profile Number                | Hits      | False Drops | Total     | Hits                      | False Drops | Total     | Extra Hits | Extra False Drops |
| 1                             | 13        | 2           | 15        | 16                        | 7           | 23        | 3          | 5                 |
| 2                             | 5         | 4           | 9         | 7                         | 4           | 11        | 2          | 0                 |
| 3                             | 4         | 0           | 4         | 4                         | 0           | 4         | 0          | 0                 |
| 4                             | 15        | 1           | 16        | 15                        | 1           | 16        | 0          | 0                 |
| 5                             | 36        | 2           | 38        | 45                        | 3           | 48        | 9          | 1                 |
| 8                             | 12        | 7           | 19        | 12                        | 7           | 19        | 0          | 0                 |
| 9                             | 23        | 5           | 28        | 25                        | 6           | 31        | 2          | 1                 |
| 10                            | 11        | 1           | 12        | 11                        | 1           | 12        | 0          | 0                 |
| 11                            | 16        | 1           | 17        | 16                        | 1           | 17        | 0          | 0                 |
| 12                            | 11        | 0           | 11        | 11                        | 0           | 11        | 0          | 0                 |
| 13                            | 4         | 0           | 4         | 4                         | 0           | 4         | 0          | 0                 |
| 15                            | <u>19</u> | <u>0</u>    | <u>19</u> | <u>19</u>                 | <u>0</u>    | <u>19</u> | <u>0</u>   | <u>0</u>          |
|                               | 169       | 23          | 192       | 185                       | 30          | 215       | 16         | 7                 |

From these data we draw the following conclusions:

1. That the addition of the broad terms does not produce an "avalanche" of retrieved references. There was a modest increase of 23 in this experiment.
2. That the ratio of hits to the total does not change significantly. The ratio was 88% before posting and 86% after.
3. That better results are obtained using the file with broad terms posted. An increase of 9% in the number of hits (16/169) was obtained.

Custom profiles are usually designed for a narrower interest than the standards tested above. Improved results should be forthcoming with the new file for custom profiles too. The searcher may choose a keyword at any level in the hierarchy and will obtain all references to more specific subject matter related to that keyword.